

УДК 339.54.012

Андреев Дмитрий Игоревич, магистрант, МИРЭА - Российский технологический университет, г. Москва

КЛАССИФИКАЦИЯ МУЗЫКАЛЬНЫХ КОМПОЗИЦИЙ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Аннотация. Классификация музыкальных композиций является одной из ключевых задач музыкальной информатики, актуальность которой связана с возрастающим количеством доступной музыки и необходимостью её систематизации и автоматической категоризации. В статье подробно рассматривается использование методов машинного обучения, а именно конволюционно-рекуррентных нейронных сетей (CRNN), для решения задач классификации музыкальных произведений по жанрам, исполнителям и инструментам. Описаны архитектурные особенности, преимущества и недостатки CRNN, приведены результаты экспериментальных исследований, подтверждающих эффективность данного подхода.

Ключевые слова: классификация музыки, машинное обучение, нейронные сети, CRNN, конволюционные сети, рекуррентные сети, музыкальная информатика.

Abstract. Classification of musical compositions is one of the key tasks of music informatics, the relevance of which is associated with the increasing amount of available music and the need for its systematization and automatic categorization. The paper discusses in detail the use of machine learning methods, namely convolutional recurrent neural networks (CRNN), to solve the problems of classifying musical works by genre, performers and instruments. The architectural features, advantages and disadvantages of CRNNs are described, and the results of experimental studies confirming the effectiveness of this approach are presented.

Keywords: music classification, machine learning, neural networks, CRNN, convolutional networks, recurrent networks, music informatics.

Введение

В современном мире объем доступной музыки постоянно растет, что делает актуальной задачу её автоматизированной обработки и классификации. Традиционные подходы, основанные на ручной разметке и экспертных оценках, становятся непрактичными в силу больших временных затрат и субъективности оценок. Поэтому в последние годы активно развивается область автоматизированного анализа музыки с использованием методов искусственного интеллекта, в особенности нейронных сетей.

Конволюционно-рекуррентные нейронные сети (CRNN) являются одним из наиболее перспективных направлений в области музыкального анализа благодаря их способности эффективно извлекать как локальные, так и глобальные признаки аудиосигналов.

Обзор существующих методов

Ранние подходы к классификации музыки использовали различные акустические признаки (темп, тональность, громкость), которые извлекались вручную или с использованием базовых алгоритмов. Среди популярных методов классификации были такие алгоритмы, как метод опорных векторов (SVM), случайные леса (Random Forest), гауссовские смеси (GMM). Однако их основным недостатком была ограниченная способность улавливать временные и пространственные взаимосвязи между признаками.

Развитие глубокого обучения существенно изменило подходы к решению задач музыкальной классификации. Нейронные сети, особенно конволюционные (CNN) и рекуррентные (RNN), позволили автоматизировать процесс выделения признаков и учитывать сложные паттерны в аудиосигналах.

Теоретические основы CRNN

Конволюционные нейронные сети (CNN) способны эффективно извлекать спектральные признаки из спектрограмм аудиосигналов. Основой их работы являются свёрточные слои, которые выявляют локальные паттерны в данных. Тем не менее, CNN недостаточно эффективно учитывают изменения этих признаков во времени, так как используют преимущественно статические методы агрегации, такие как пулинг (pooling).

В отличие от них, рекуррентные нейронные сети (RNN), включая LSTM и GRU, специально предназначены для обработки последовательностей и способны запоминать и учитывать контекстные временные зависимости, что критически важно при анализе музыкальных композиций.

Архитектура и преимущества CRNN

Конволюционно-рекуррентные сети сочетают преимущества CNN и RNN в единой архитектуре. CNN выполняют первичную обработку аудиоданных, выделяя спектральные и пространственные признаки, после чего рекуррентные слои анализируют последовательности этих признаков во времени, захватывая глобальные паттерны.

Архитектура CRNN обычно состоит из нескольких этапов:

1. Конволюционные слои – извлекают локальные признаки и уменьшают размерность входных данных.
2. Рекуррентные слои (GRU или LSTM) – учитывают временные зависимости и глобальные изменения признаков.
3. Полносвязный слой – осуществляет окончательную классификацию по заданным категориям.

Экспериментальная оценка CRNN

Экспериментальная проверка CRNN была проведена на наборе данных Million Song Dataset с метками, предоставленными сервисом Last.fm. В рамках исследований использовались логарифмические мел-спектрограммы аудиосигналов. Полученные результаты подтвердили, что архитектура CRNN демонстрирует превосходство над традиционными конволюционными нейронными сетями по ряду критериев, особенно по метрике AUC-ROC. Это связано с тем, что рекуррентные слои сети позволяют более эффективно агрегировать временные признаки, что критически важно для анализа музыкальных данных.

Также важным экспериментом стала проверка способности CRNN к идентификации исполнителей на наборе данных Artist20. В этой задаче нейросеть смогла достигнуть среднего значения F1-меры до 93,7%, что существенно превысило результаты ранее используемых классических подходов, таких как метод опорных векторов (SVM) и гауссовские смеси (GMM). Высокий уровень точности объясняется способностью CRNN выделять и анализировать уникальные стилистические характеристики каждого исполнителя, учитывая как локальные спектральные, так и глобальные временные признаки.

Проблемы и ограничения CRNN

Несмотря на высокие показатели эффективности, у CRNN есть ряд существенных ограничений, которые необходимо учитывать при их применении:

1. **Вычислительная сложность.**

Глубокие нейронные сети требуют значительных вычислительных ресурсов и времени для обучения, что ограничивает их использование в некоторых практических задачах.

2. **Необходимость больших объемов данных.**

Для достижения высоких результатов и предотвращения переобучения необходимо иметь доступ к большому количеству размеченных аудиоданных.

3. Трудности в распознавании сходных стилей.

CRNN могут испытывать затруднения при работе с музыкальными композициями исполнителей, имеющих сходные стилистические особенности, что требует дополнительных методик и подходов для повышения дискриминации.

Направления будущих исследований

Развитие CRNN связано с несколькими перспективными направлениями:

1. Использование механизмов внимания (attention). Введение механизма внимания позволяет модели более целенаправленно выделять значимые области спектрограммы и эффективнее учитывать контекстные зависимости во времени.

2. Улучшение предварительной обработки аудиоданных. Разработка и применение новых техник предварительной обработки данных, таких как аугментация (изменение темпа, тональности, добавление шума и т.д.), могут повысить устойчивость модели и её способность к обобщению.

3. Интеграция данных разных модальностей. Комбинирование аудио с текстовыми описаниями, обложками альбомов или визуальными аспектами музыкальных клипов может дать дополнительный прирост точности классификации и расширить область применения нейросетей.

4. Адаптация архитектуры под другие музыкальные задачи, например, автоматическую транскрипцию или распознавание эмоциональной окраски, позволит значительно расширить спектр задач, решаемых с помощью CRNN.

ЗАКЛЮЧЕНИЕ

Использование конволюционно-рекуррентных нейронных сетей для задач классификации музыкальных композиций демонстрирует высокие результаты и значительные преимущества по сравнению с традиционными алгоритмами машинного обучения. Гибридная архитектура CRNN, сочетающая локальный спектральный анализ конволюционными слоями и глобальный временной анализ рекуррентными слоями, позволяет эффективно извлекать и использовать комплексные структуры данных аудиосигналов.

Будущие исследования и разработки должны быть направлены на оптимизацию архитектурных решений, интеграцию новых подходов и методов обработки данных, а также расширение применимости CRNN на более сложные задачи анализа и классификации музыки. Это позволит улучшить точность, сократить время обучения и расширить спектр прикладных задач, успешно решаемых с использованием методов машинного обучения и глубоких нейронных сетей.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Choi K., Fazekas G., Sandler M., Cho K. Convolutional recurrent neural networks for music classification // ICASSP, 2017.
2. Dieleman S., Schrauwen B. End-to-end learning for music audio // IEEE ICASSP, 2014.
3. Sigtia S., Dixon S. Improved music feature learning with deep neural networks // IEEE ICASSP, 2014.
4. Nasrullah Z., Zhao Y. Music artist classification with convolutional recurrent neural networks // arXiv preprint arXiv:1901.04555, 2019.
5. Chiliguano P., Fazekas G. Hybrid music recommender using content-based and social information // IEEE ICASSP, 2016.
6. McFee B., Raffel C., Liang D., Ellis D. P., McVicar M., Battenberg E., Nieto O. librosa: Audio and music signal analysis in python // Proceedings of the 14th python in science conference, 2015.
7. Mandel M., Ellis D. Song-level features and support vector machines for music classification // ISMIR, 2005.
8. Whitman B., Flake G., Lawrence S. Artist detection in music with minnowmatch // IEEE Signal Processing Society Workshop, 2001.